

Intern Homework README

- 問題
 - 給予文章發佈後前 6 小時的狀態及基本資訊，預測 24 小時後的愛心數
 - Evaluation
 - Mean Absolute Percentage Error ([wiki](#))
 - Dataset
 - Feature
 - `title`: 文章標題
 - `created_at`: 文章發佈時間
 - `like_count_1~6h`: 文章發佈後 1~6 小時的累積愛心數
 - `comment_count_1~6h`: 文章發佈後 1~6 小時的累積留言數
 - `forum_id`: 文章發佈看板 ID
 - `author_id`: 文章作者 ID
 - `forum_stats`: 看板資訊
 - Label
 - `like_count_24h`: 文章發布後 24 小時的累積愛心數
 - 附件內容
 - https://drive.google.com/drive/folders/1wb_BDqUTzl80cg_LYJ6CYBf5J5ZxnZDk
 - Train Set: `intern_homework_train_dataset.csv` 共 50000 筆資料
 - Public Test Set: `intern_homework_public_test_dataset.csv` 共 10000 筆資料
 - Private Test Set: `intern_homework_private_test_dataset.csv` 共 10000 筆資料
 - Example Result: `example_result.csv`
 - 繳交方式
 - 將以下項目打包壓縮後透寄至 dcard.internprogram@dcard.cc 繳交，請參考以下注意事項：
 - ```
(your_name)_2023_dcard_ml_intern_homework
├── report.pdf
├── result.csv
└── src
 └── README.md
```
      - **Filename** : `(your_name)_2023_dcard_ml_intern_homework.zip`
      - Report: 完整的報告 (pdf) 描述使用的演算法、實驗結果以及想法
      - Code: 可執行的 python code，需附上 README 描述如何使用
      - Result: 將 Private Test Set 的預測結果 (依照原本 dataset 的順序) 存成 `result.csv`，請參考附件 `example_result.csv`
  - 評鑑方法
    - Evaluate 作業的預測結果後，將所有面試者的結果排名
    - 依據報告內容的完整度評分
    - 依據 Code Quality 評分
    - 綜合上述因素選出表現最好的人選進行下階段面談
-

# Intern Homework README (EN)

- Problem
  - Given the status and basic information within the first 6 hours after the post is published, predict the number of likes after 24 hours.
- Evaluation
  - Mean Absolute Percentage Error ([wiki](#))
- Dataset
  - Feature
    - `title`: Title of the post
    - `created_at`: Publish time of the post
    - `like_count_1~6h`: The cumulative number of likes within 1-6 hours after the post is published
    - `comment_count_1~6h`: The cumulative number of comments within 1-6 hours after the post is published
    - `forum_id`: Forum ID of the post
    - `author_id`: Author ID of the post
    - `forum_stats`: Forum stats of the post
  - Label
    - `like_count_24h`: The cumulative number of likes 24 hours after the post is published
- Attachment
  - [https://drive.google.com/drive/folders/1wb\\_BDqUTzl80cg\\_LYJ6CYBf5J5ZxnZDk](https://drive.google.com/drive/folders/1wb_BDqUTzl80cg_LYJ6CYBf5J5ZxnZDk)
  - Train Set: `intern_homework_train_dataset.csv` Total 50000 data points
  - Public Test Set: `intern_homework_public_test_dataset.csv` Total 10000 data points
  - Private Test Set: `intern_homework_private_test_dataset.csv` Total 10000 data points
  - Example Result: `example_result.csv`
- Submission
  - Please compress the following items into a package and send them to [dcard.internprogram@dcard.cc](mailto:dcard.internprogram@dcard.cc) for submission. Please refer to the following precautions
    - ```
(your_name)_2023_dcard_ml_intern_homework
├─ report.pdf
├─ result.csv
└─ src
   └─ README.md
```
 - **Filename** : `(your_name)_2023_dcard_ml_intern_homework.zip`
 - Report: A complete report (PDF) describing the algorithm used, experimental results, and ideas.
 - Code: Executable Python code, with a README describing how to use it.
 - Result: Save the predicted results of the Private Test Set (in the order of the original dataset) as `result.csv`. Please refer to the attached `example_result.csv`.
- Evaluation method
 - Evaluate the predicted results of the assignment and rank all interviewees' results.
 - Score based on the completeness of the report content.
 - Score based on code quality.
 - Select the best candidate based on the above factors for the next interview.